

Research Article

MHC I Stabilizing Potential of Computer-Designed Octapeptides

Joanna M. Wisniewska,¹ Natalie Jäger,¹ Anja Freier,² Florian O. Losch,²
Karl-Heinz Wiesmüller,³ Peter Walden,² Paul Wrede,⁴ Gisbert Schneider,^{1,5} and Jan A. Hiss^{1,5}

¹Institute of Organic Chemistry and Chemical Biology, Johann Wolfgang Goethe-Universität,
Siesmayerstraße. 70, 60323 Frankfurt am Main, Germany

²Clinical Research Group Tumor Immunologie, Skin Cancer Center Charité, Clinic of Dermatology,
Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

³Institut für Organische Chemie, Universität Tübingen, Auf der Morgenstelle 18, 74076 Tübingen, Germany

⁴Institute of Molecular Biology and Bioinformatics, Charité-Universitätsmedizin Berlin, Arnimallee 22, 14195 Berlin, Germany

⁵Institute of Pharmaceutical Sciences, ETH Zürich, Wolfgang-Pauli-Street. 10, 8093 Zürich, Switzerland

Correspondence should be addressed to Jan A. Hiss, jan.hiss@pharma.ethz.ch

Received 28 September 2009; Revised 27 January 2010; Accepted 8 March 2010

Academic Editor: Yongqun Oliver He

Copyright © 2010 Joanna M. Wisniewska et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Experimental results are presented for 180 *in silico* designed octapeptide sequences and their stabilizing effects on the major histocompatibility class I molecule H-2K^b. Peptide sequence design was accomplished by a combination of an ant colony optimization algorithm with artificial neural network classifiers. Experimental tests yielded nine H-2K^b stabilizing and 171 nonstabilizing peptides. 28 among the nonstabilizing octapeptides contain canonical motif residues known to be favorable for MHC I stabilization. For characterization of the area covered by stabilizing and non-stabilizing octapeptides in sequence space, we visualized the distribution of 100,603 octapeptides using a self-organizing map. The experimental results present evidence that the canonical sequence motives of the SYFPEITHI database on their own are insufficient for predicting MHC I protein stabilization.

1. Introduction

Cell surface presentation of peptides by major histocompatibility complex I (MHC I) is prerequisite for the initiation of an adaptive immune response [1] and knowledge of MHC-binding peptides is required for the development of vaccines and immunomonitoring protocols for cell-mediated immunity. MHC I molecules are integral membrane proteins that bind peptides with a length of eight up to thirteen amino acids for presentation to CD8⁺ T lymphocytes [2, 3]. Peptide binding to MHC I stabilizes the MHC-peptide structure at the cell surface of antigen presenting cells. Binding of an octapeptide to an MHC I molecule is defined by the recognition of the peptide by the MHC molecule and its binding affinity [4]. In consequence the binding of the octapeptide leads to stabilization of the MHC-peptide complex on the cell surface. Complex stability is critically influenced by the amino acid sequence of the

bound peptide [5], for which Rammensee and coworkers suggested allele-specific canonical sequence motifs [3]. For the octapeptides presented by the mouse MHC I H-2K^b this sequence motif (the canonical or SYFPEITHI motif) is defined as X-X-(Y)-X-[Y/F]-X-X-[L, M, I, V]. Positions three, five, and eight are also referred to as “anchor positions” [6].

For characterization of the H-2K^b stabilizing and nonstabilizing sequence space we designed a diverse set of octapeptides. To explore extensions and alternatives to the known canonical motif, the set of designed octapeptides included sequences containing the full, partial, or no canonical motif. To generate new octapeptides that stabilize H-2K^b we applied an Ant Colony Optimization (ACO) [7] algorithm in combination with neural network classifiers. Artificial neural networks (ANNs) [8] were trained using a set of 423 octapeptides with known H-2K^b stabilizing effect as determined in cellular stabilization assays [9]. The resulting

machine learning classifiers served as fitness function for the ACO algorithm. Navigation through sequence space containing 20^8 possible octapeptides was realized by the ACO *meta*-heuristic which is deduced from social insect behavior [7, 10]. ACO is a probabilistic technique that is not susceptible to dominant ultimate solutions but, due to its “swarm intelligence” based on numerous autonomous agents, open for broad and distributed optimization [11]. New peptide sequences were generated with the ACO algorithm and presented to the trained ANNs for fitness evaluation. During this optimization process the peptide sequences were iteratively adapted according to the ANN fitness score. Finally, the designed octapeptides were synthesized and their stabilization effect was tested experimentally.

We present evidence that rational peptide design utilizing ACO is feasible and leads to novel bioactive peptides with minimal experimental effort. Here, the focus is on the *de novo* design of peptides with a specific MHC I stabilization effect. While some of the designed peptides conform to the known canonical motif for H-2K^b stabilizing peptides, we also show that the degree to which the peptide sequence matches the motif alone is insufficient for prediction of MHC I stabilization. We designed peptides with the complete canonical sequence motif but lacking detectable stabilizing effect. For visualization of the transition between stabilizing and nonstabilizing octapeptides we present a projection of peptide sequence space on a self-organizing map (SOM). This form of representation facilitates the identification of clusters of stabilizing peptides based on their physicochemical properties.

2. Materials and Methods

2.1. Data Set. Training data were compiled from the public databases (AntiJen [12], EPIMHC [13], IEDB [14], MHCBN [15]) and literature sources [16, 17]. The complete dataset contained 423 octapeptides with 242 positive (stabilizing) and 181 negative (nonstabilizing) examples. The annotation of octapeptides as stabilizing and nonstabilizing mouse MHC I protein H-2K^b was based on published experimental data. EC₅₀ values below 10 μM were regarded as H-2K^b stabilizing, greater EC₅₀ values as nonstabilizing.

2.2. Sequence Encoding. Each residue of an octapeptide was encoded by five different sets of molecular descriptors (See supplementary material (Suppl. 1) available online at doi:10.1155/2010/396847). The combination of amino acid descriptors served as input for the ANNs. The dimension of the input originated from the coding of each amino acid of the octapeptide by each descriptor.

2.3. The Ant Colony Optimization. The ACO algorithm was implemented using the Java programming language V.1.6 (Sun Microsystems, Inc., Santa Clara, CA, USA). Our ACO algorithm is defined by three consecutive steps: sequence design, path evaluation, and pheromone update, as previously described by Jäger et al. [18]. Peptide design by ACO was terminated when the pheromone concentration

had been constant for 10,000 iterations. Together the three steps represent a single iteration of the algorithm (one generation of ants). Ants are computational agents with individual memory coded via “pheromone concentrations”. While moving through the search space each ant generates a path corresponding to a new octapeptide. All ants of one generation move independent of each other on individual paths. The resulting paths were evaluated by a fitness function implemented as ANNs. Communication between subsequent generations of ants is achieved through the modification of pheromone concentrations (“stigmergy” [19, 20]). The pheromone matrix represents the collective memory of an ant colony. Only the path with the highest fitness obtained a pheromone update. The advantage of the ACO algorithm is that agents need no information about the complete problem to propose a solution, in our case the complete possible sequence space containing 20^8 octapeptides.

2.4. Artificial Neural Network Fitness Function. Fully connected feedforward networks with a single hidden layer and one output neuron (all neurons with sigmoidal activation) were implemented using Matlab (version 7.4.0.287 R2007a, The Mathworks Inc.; neural networks toolbox version 5.0.2). The outputs of five ANNs were combined as input for a jury network [21]. The output of the jury served as fitness value (or “score”) for the ACO algorithm, which adopted values of the interval]0, 1[. Details on the network architecture were described previously [17].

2.5. Stabilization Assay. The stabilization assay was performed as described by Brock et al. [9] using TAP-deficient RMA-S cells (mutagenized Rauscher virus-induced T lymphoma cells of mouse origin) [22]. The cells were cultured in DMEM (Gibco-BRL, Karlsruhe, Germany) with 10% FCS (Sigma-Aldrich, Steinheim, Germany) at 37°C with 8% CO₂. For accumulation of peptide-free MHC I proteins at the cell surface, the cells were cultured for 16 hours at 26°C. The cells were incubated with the peptides in 10 serial dilutions of 100 to 5.6×10^{-4} μg/mL at room temperature for 1 hour, followed by 1-hour incubation at 37°C for denaturation of peptide-free MHC I proteins. The stabilized MHC I proteins were visualized and quantified by flow cytometry using the H-2K^b specific monoclonal antibody B8.24.3 [23] purified in the laboratory from hybridoma culture supernatant by protein G affinity chromatography (Pierce, Darmstadt, Germany) and an R-Phycoerythrin-conjugated anti-mouse antibody (Dianova GmbH, Hamburg, Germany) as secondary reagent. The stabilizing effect of the peptides was determined as mean fluorescence intensity (MFI). The EC₅₀ value is the peptide concentration that is required for half-maximal stabilization of the MHC I molecules at the cell surfaces (half-maximal MFI). All peptides were custom-synthesized by EMC microcollections GmbH (Tübingen, Germany).

2.6. SYFPEITHI Score (S-Score). The S-score was calculated using the public web server at URL: <http://www.syfpeithi.de/>

(version July 2009). The *S*-score [6] indicates how well a peptide sequence matches the canonical motif.

2.7. IEDB-ANN Score. For calculation of the Immune Epitope Database- (IEDB-)ANN score the public web server at URL: http://tools.immuneepitope.org/analyze/html/mhc_binding.html (version 2009-09-01) was used. IEDB offers several prediction tools for peptide binding to MHC I molecules (artificial neural networks (ANNs), average relative binding (ARB), stabilized matrix method (SMM), SMM with a peptide: MHC binding energy covariance matrix (SMMPMBEC), scoring matrices derived from combinatorial peptide libraries (complib_sidney2008), consensus) [24]. The IEDB-ANN method [25] was chosen because it has been determined to be qualitatively best performing [26]. The IEDB-ANN scores are predicted IC_{50} values.

2.8. Self-Organizing Map (SOM). For visualization of the peptide distribution in a high-dimensional descriptor space we used planar SOMs [27] as implemented in the molmap software package [28, 29]. The trained SOM performs a nonlinear mapping from the original descriptor space onto a two-dimensional map. Each data point is assigned to one of a defined number of receptive fields (neurons) of the SOM. SOM training was performed as described previously [30].

3. Results and Discussion

We report the design and examination of 180 octapeptides (Table 1; Suppl. 2) in a cellular MHC I stabilization assay. The ability of an octapeptide to stabilize MHC I was specified as EC_{50} value, which is defined as the peptide concentration required for half-maximal stabilization of the MHC I proteins at the cell surface by the test peptide. Nine of the *in silico* designed octapeptides exhibited a stabilizing effect (Table 1, Seq. 1–9), and 171 were nonstabilizing (Table 1, Seq. 10–50; Suppl. 2, Seq. 51–180). Six of the nine stabilizing octapeptides had EC_{50} values below $10\ \mu\text{M}$ (Table 1, Seq. 4–9) (i.e., strong MHC I stabilization). Three octapeptides (Table 1, Seq. 1–3) can be regarded as medium stabilizers ($EC_{50} > 20\ \mu\text{M}$), two of which completely matched the canonical motif (Table 1, Seq. 1 and 2) with EC_{50} values of $24\ \mu\text{M}$ and $20\ \mu\text{M}$. Six of the nine octapeptides that correspond to the canonical motif in only two of the three anchor positions yielded EC_{50} values below $10\ \mu\text{M}$. Peptide 3 had an EC_{50} of $25\ \mu\text{M}$. Peptide 8 (WKFIFDPV) conforming to the SYFPEITHI motif in two positions (underlined) was the most potent peptide with an EC_{50} of $0.4\ \mu\text{M}$. Peptide 9 (FHHAHRTV) obeys the canonical motif in just one anchor position but was still among the best stabilizers with an EC_{50} value of $9\ \mu\text{M}$.

The SYFPEITHI score (*S*-score) is used as a computed index for prediction of stabilizing abilities of peptides for specific MHC molecules [6]. A high value indicates strong stabilizing effects. The *S*-score for the positive control in our experiments (SIINFEKL from ovalbumin [4]) is 25. The *S*-score of a known nonstabilizing octapeptide (LSPFPFDL, an endogenous MHC I H2-L^d epitope [31]) is 13. The

computed *S*-scores for the nine stabilizing octapeptides were between 8 and 27 (mean = 20 ± 6) reflecting their stabilizing effect (outlier: octapeptide 9 with an *S*-score of 8). Peptides 4, 5, and 7, while exhibiting EC_{50} values similar to sequence 9 ($EC_{50} = 9\ \mu\text{M}$), have more than two times greater *S*-scores (*S*-score = 22, *S*-score = 20, *S*-score = 17 (Table 1, Seq. 4, 5, 7)). A possible explanation for the deviation between the SYFPEITHI score and the actual binding behavior could be the anchor position assignment. Sequences 4, 5, and 7 completely fulfill the canonical motif while sequence 9 fulfills it in only one position. Thus the degree of correspondence to the canonical motif is well represented by the *S*-score but does not necessarily reflect the actual binding behavior. This suggests that alternative sequence motifs might confer strong stabilization effects or that the binding motive concept needs to be extended.

We then compared our experimental results to predictions of the Immune Epitope Database (IEDB) [24]. The database offers several prediction methods of which, according to Peters et al. [26], IEDB-ANN [25] is the best performing. For the nine binding peptides found by us, the Pearson correlation [32] between the IC_{50} values predicted by IEDB-ANN for mouse H2-K^b and our measured EC_{50} values is -0.34 , which indicates moderate negative correlation. Using the activity cutoff of $IC_{50} < 500\ \text{nM}$ for “medium activity” [25], the IEDB-ANN method correctly predicts four of nine sequences as binding peptides (Table 1, Seq. 1–3, 6).

The remaining 171 octapeptides showed no detectable stabilizing effect at a maximal experimental peptide concentration of $100\ \mu\text{g}/\text{mL}$ and were therefore defined as nonstabilizing (Table 1, Seq. 10–50; cf. Suppl. 2, Seq. 51–180). The nonstabilizing octapeptides can be grouped into four categories according to the degree of fulfillment of the canonical SYFPEITHI motif:

Category (i): three canonical anchor amino acids: 12 octapeptides (Table 1, Seq. 10–21),

Category (ii): two canonical anchor amino acids: 16 octapeptides (Table 1, Seq. 22–37),

Category (iii): one canonical anchor amino acid: 23 octapeptides (Table 1, Seq. 38–42; Suppl. 2, Seq. 50–68),

Category (iv): no canonical anchor amino acids: 120 octapeptides (Table 1, Seq. 42–50; Suppl. 2, Seq. 69–180).

For octapeptides of category (i) high *S*-scores were computed in the range between 22 and 28 (mean = 25 ± 2) suggesting a stabilizing ability of the octapeptides. In comparison to the *S*-scores of the nine stabilizing octapeptides, category (i) sequences had higher *S*-scores thus erroneously predicting an even stronger MHC I stabilizing effect. Category (ii) peptides obtained a mean *S*-score of 19 ± 2 still indicating possible MHC I stabilization. Notably, none of these octapeptides had a stabilizing effect in our experiments. The *S*-scores of category (iii) peptides (mean = 11 ± 2) are in agreement with the lack of a stabilizing effect. For category (iv) peptides the computed *S*-scores (mean = 1 ± 1) perfectly agreed with the experimental results obtained for these 120 sequences.

TABLE 1: Experimentally tested octapeptides. Sequences are given in single-letter code; motif anchor positions fulfilling the SYFPEITHI motif are underlined. S-score: SYFPEITHI score [6]. IEDB-ANN score: predicted IC₅₀ values by IEDB-ANN method [25]. <500 nM: recommended cutoff for IEDB-ANN score [25] for binders predicted with intermediate affinity; *active*: IEDB-ANN score <500 nM; 0: IEDB-ANN score >500 nM. EC₅₀ values correspond to the peptide concentration required for 50% of maximal MHC I protein stabilization; values in brackets are standard deviations ($N = 3$). n.d.: not detectable.

	Peptide number	Sequence	S-score	IEDB-ANN score [μ M]	<500 nM	Experimental EC ₅₀ [μ M]
<i>Stabilizing peptides</i>	1	FRYP <u>Y</u> KTL	27	0.5	active	24.0 (\pm 11)
	2	FRY <u>I</u> YHTL	27	0.2	active	20.0 (\pm 4.0)
	3	FHW <u>D</u> YRGL	22	0.4	active	25.0 (\pm 11)
	4	WRFK <u>Y</u> DNL	22	2.3	0	7.0 (\pm 4.0)
	5	WRFV <u>Y</u> WRL	20	1.2	0	8.0 (\pm 4.0)
	6	WRF <u>I</u> YFNL	21	0.4	active	3.6 (\pm 1.2)
	7	WDFK <u>F</u> DSV	17	4.7	0	9.0 (\pm 4.0)
	8	WKF <u>I</u> FDPV	16	3.0	0	0.4 (\pm 0.1)
	9	FHH <u>A</u> HRTV	8	10.5	0	9.0 (\pm 4.0)
<i>Nonstabilizing peptides of category (i)</i>	10	FRY <u>E</u> YRSL	28	0.1	active	n.d.
	11	WR <u>Y</u> IYHSI	22	2.5	0	n.d.
	12	HR <u>Y</u> VYRNI	24	0.4	active	n.d.
	13	FH <u>Y</u> AYRSV	23	0.1	active	n.d.
	14	YR <u>Y</u> KYDRL	27	0.5	0	n.d.
	15	WR <u>Y</u> QYDNL	27	0.7	0	n.d.
	16	WR <u>Y</u> R <u>Y</u> WSL	27	0.6	0	n.d.
	17	WR <u>Y</u> NYDPL	26	1.7	0	n.d.
	18	WR <u>Y</u> HYDPL	26	2.4	0	n.d.
	19	WK <u>Y</u> QYDNL	27	0.3	active	n.d.
	20	WK <u>Y</u> I <u>F</u> DPV	22	1.2	0	n.d.
21	WK <u>Y</u> PFDPV	22	2.0	0	n.d.	
<i>Nonstabilizing peptides of category (ii)</i>	22	FR <u>Y</u> LYKNA	17	1.2	0	n.d.
	23	FR <u>Y</u> VWRTL	18	1.7	0	n.d.
	24	FH <u>Y</u> LYHTA	17	0.9	0	n.d.
	25	FR <u>Y</u> PYHTP	17	17.8	0	n.d.
	26	FR <u>W</u> EYRGL	22	0.7	0	n.d.
	27	FR <u>H</u> IYRTI	18	9.3	0	n.d.
	28	FR <u>H</u> GYRQI	18	12.6	0	n.d.
	29	FH <u>W</u> AYHTV	17	1.2	0	n.d.
	30	WR <u>W</u> LYKGV	16	4.3	0	n.d.
	31	WR <u>F</u> PYDQL	21	6.6	0	n.d.
	32	WRFK <u>Y</u> DPL	21	3.3	0	n.d.
	33	WR <u>F</u> PYDKL	21	6.8	0	n.d.
	34	WRFV <u>Y</u> DNL	21	1.9	0	n.d.
	35	WKF <u>K</u> FDPV	17	2.8	0	n.d.
	36	WK <u>I</u> NFDPV	17	8.2	0	n.d.
	37	TTEW <u>Y</u> TKI	18	3.4	0	n.d.
<i>5 of 23 nonstabilizing peptides of category (iii)</i>	38	RGEV <u>F</u> TAT	13	24.0	0	n.d.
	39	FH <u>Y</u> DHRNA	9	6.5	0	n.d.
	40	HR <u>W</u> FWQP	11	26.0	0	n.d.
	41	SIK <u>N</u> FHYH	12	13.6	0	n.d.
	42	FR <u>H</u> DYHSP	11	29.2	0	n.d.

TABLE 1: Continued.

Peptide number	Sequence	S-score	IEDB-ANN score [μM]	<500 nM	Experimental EC ₅₀ [μM]
43	TVEQGVTQ	1	37.5	0	n.d.
44	FHHGHNVP	0	36.8	0	n.d.
45	QDGHEIHR	1	38.4	0	n.d.
46	TVEQGVTQ	1	37.5	0	n.d.
47	GVDQSYLK	0	38.1	0	n.d.
48	SVENPILR	2	33.7	0	n.d.
49	NEGWTIHR	1	38.8	0	n.d.
50	SVDHSIFK	2	36.1	0	n.d.

For category (iii) 5 of 23 octapeptides are given as examples. The remaining 18 octapeptides of category (iii) are listed in Suppl. 2 (Seq. No. 51–68). For category (iv) 7 of 120 octapeptides are given as examples. The remaining 113 octapeptides of category (iv) are listed in Suppl. 2 (Seq. No. 69–180).

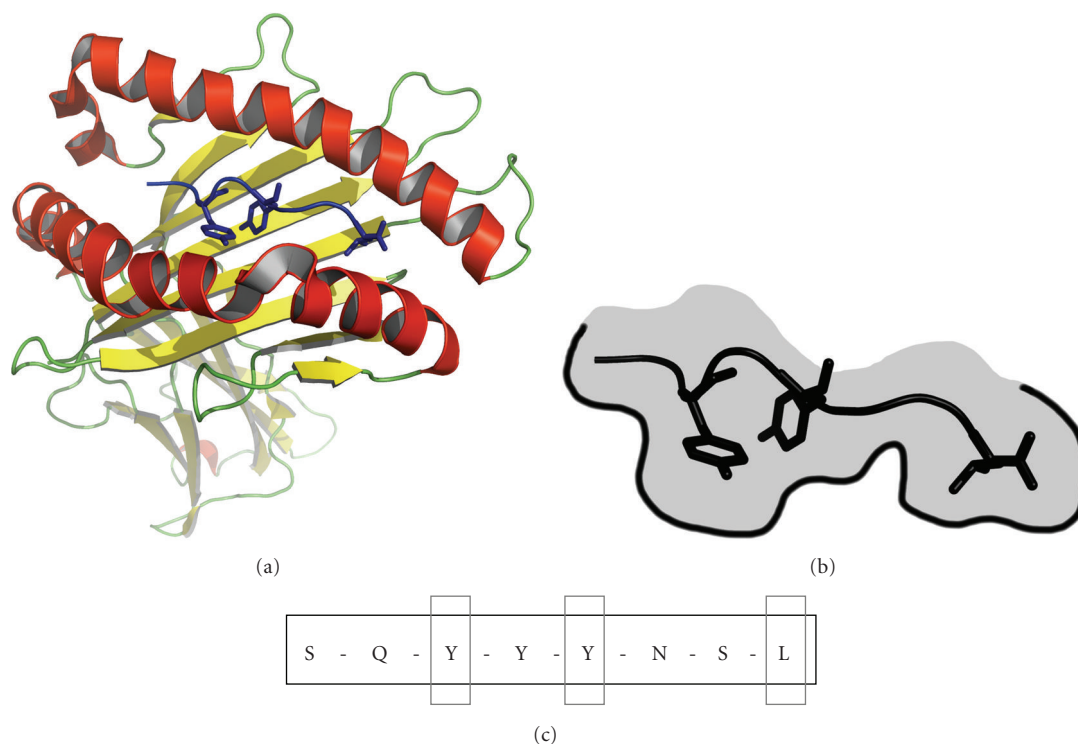


FIGURE 1: Structural model of murine MHC I molecule H2-K^b with bound octapeptide SQYYNSL (PDB entry 2clv [33]). (a) Crystal structure model of H2-K^b with bound octapeptide SQYYNSL (PDB entry 2clv [33]). Top view of the peptide binding groove: *red*: alpha helices; *yellow*: beta sheets; *blue*: octapeptide SQYYNSL, amino acids of the anchor positions are shown as sticks. (b) Schematic side view of the peptide binding groove with the bound octapeptide SQYYNSL; Amino acid side chains at the anchor positions are shown as sticks. (c) The octapeptide SQYYNSL with boxed anchor positions of the canonical motif.

The IEDB-ANN method [25] predicts four sequences as “binding”, which were determined as “nonbinding” in our experiments (Table 1, category (ii), Seq. 10, 12, 13, 19). The remaining 37 negative sequences are correctly predicted as “nonbinding” (Table 1, categories (ii)–(iv)). Compared to the S-score index, the IEDB-ANN method is better suited for identifying nonbinding sequences that contain only a partial canonical motif (categories (ii) and (iii)). Despite these differences, both software tools (S-score and IEDB-ANN method) can be recommended for identification of negative

(inactive) sequences lacking the canonical motif (categories (iii)–(iv)). Based on these limited data, quantitative IC₅₀ predictions of binding/nonbinding peptides by this software seem to be of limited accuracy but qualitative prediction is acceptable.

The experimental results for the 180 designed octapeptides allowed us to reassess the canonical motif. We found 28 inactive octapeptides that conform to the motif in all three (category (i)) or two residue positions (category (ii)). This corroborates the results of Zhong et al. [16] reporting

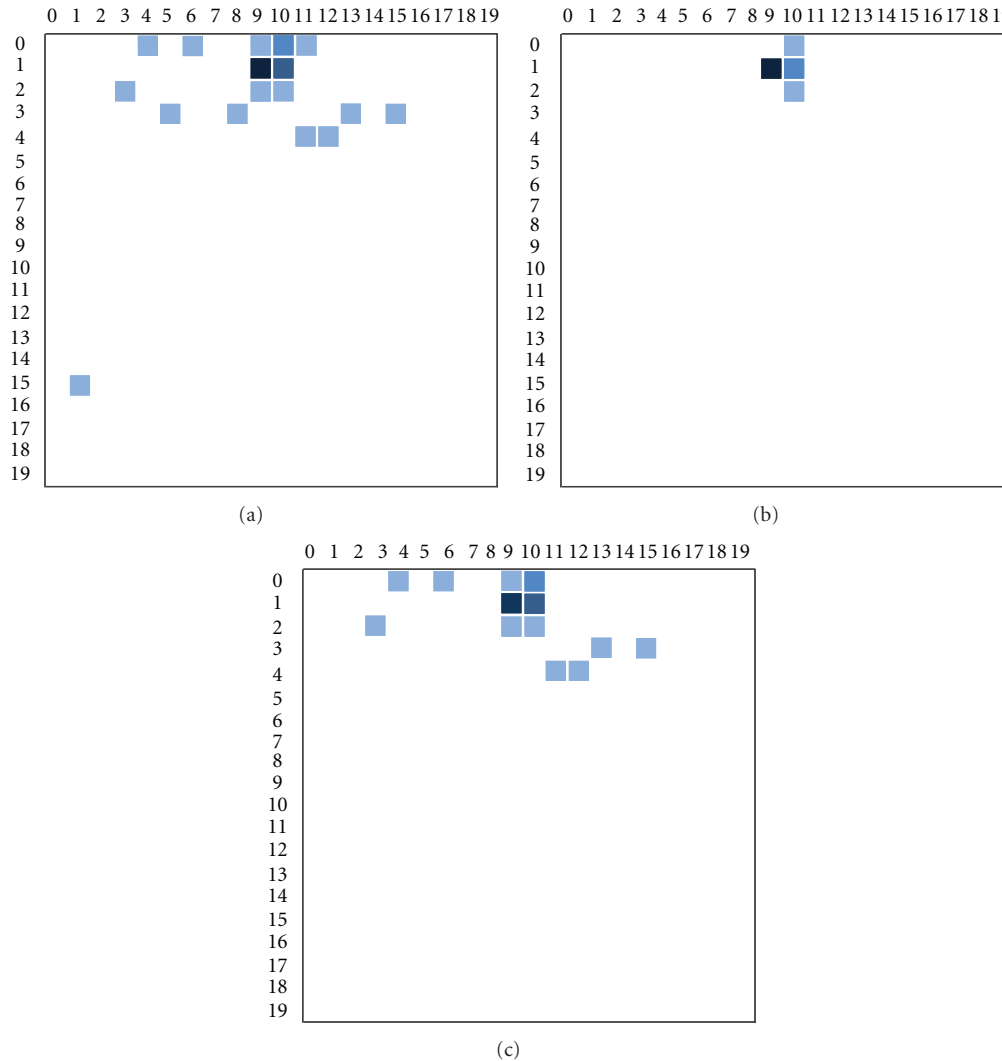


FIGURE 2: Self-organizing map (SOM) projection of MHC I H2-K^b stabilizing peptides. The SOM was trained with 100,603 octapeptides (603 peptides with known H-2K^b binding affinity and 100,000 octamer sequences randomly generated according to the amino acid frequency of *M. musculus* proteins). Only the distribution of the known stabilizing peptides is shown. The color of a neuron reflects its relative occupancy (white: empty; light blue: slightly occupied; dark blue: most occupied). (a) Distribution of the known 251 H-2K^b stabilizing octapeptides. (b) Distribution of 80 stabilizing octapeptides with amino acid sequences corresponding to the canonical motif in all three anchor positions. (c) Distribution of 152 stabilizing octapeptides with amino acid sequences corresponding to the canonical motif in two residue positions.

one nonstabilizing octapeptide with the canonical motif, and Hiss et al. [17] reporting four nonstabilizing octapeptides corresponding to the motif. Our data suggest that the canonical motif alone is insufficient for predicting MHC I stabilization. Octapeptides stabilize the MHC I molecules by binding into the peptide binding groove which is framed by two alpha helices on top of a eight-stranded beta sheet [2] (Figure 1(a)). Amino acids at sequence position three and five, favorably tyrosine or phenylalanine, can form aromatic interactions with MHC I residues facing the binding groove (Figure 1(b)), which could explain why canonical occupancy often corresponds to stabilizing peptides [34]. In addition, the aliphatic residue at position eight interacts with aliphatic amino acids in a deep pocket of the MHC I peptide binding canyon (Figure 1(b)). Octapeptides that conform completely

to the canonical motif but show no stabilizing effect indicate that other amino acids besides the three anchor residues are important for the stabilizing effect. Amino acids at nonmotif positions could interfere with the favorable effects of the three anchor residues and lead to a nonstabilizing peptide.

To visualize the distributions of stabilizing and nonstabilizing octapeptides, we trained a SOM [28, 29] to obtain a two-dimensional map of the peptide distribution. The SOM represents the peptides based on their physicochemical properties coded by a multidimensional vector. Adjacent regions of a given peptide on the SOM represent peptides with similar physicochemical properties. The SOM was trained with a total of 100,603 octapeptides. We randomly generated 100,000 octamer sequences according to the amino acid frequency found in known mouse proteins to mimic

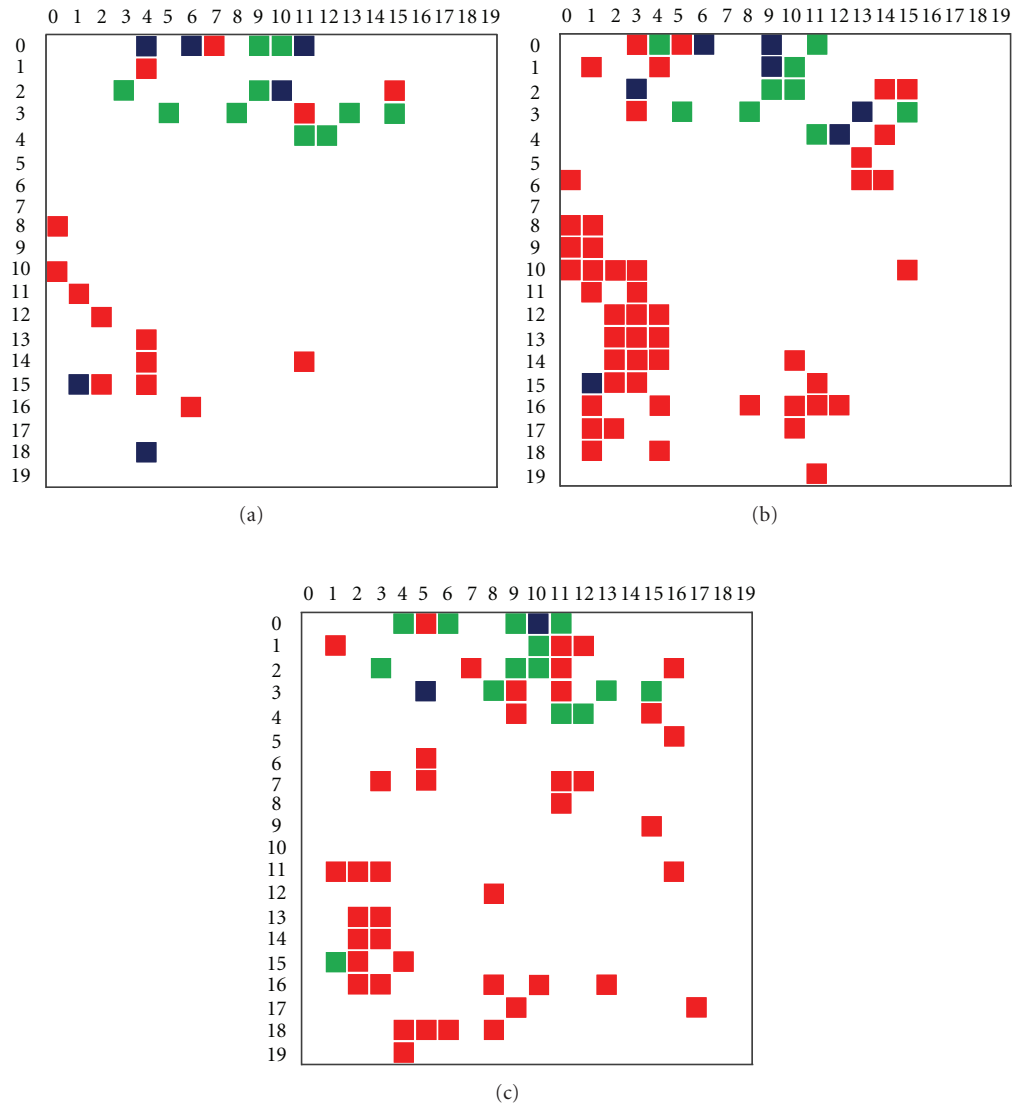


FIGURE 3: Self-organizing map (SOM) projection of MHC I stabilizing peptides. The SOM was trained with 100,603 octapeptides (603 peptides with known H-2K^b binding affinity and 100,000 octamer sequences randomly designed according to the amino acid frequency of *M. musculus* proteins). Neurons are colored according to the MHC I stabilizing ability of the octapeptides (red: nonstabilizing; blue: mixed; green: stabilizing). (a) Projection of the known 251 stabilizing octapeptides and 24 nonstabilizing octapeptides with amino acid sequences corresponding to the canonical motif in all three positions. (b) Projection of the known 251 stabilizing octapeptides and 93 nonstabilizing octapeptides with amino acid sequences corresponding to the canonical motif in two positions. (c) Projection of the known 251 stabilizing octapeptides and 59 nonstabilizing octapeptides with amino acid sequences corresponding to the canonical motif in only one position.

murine sequence space. For the remaining 603 octapeptides the H-2K^b binding affinities were known from published experimental data (training data set: 423 octapeptides with 242 stabilizing and 181 nonstabilizing peptides and own experimental results: 180 octapeptides with 9 stabilizing and 171 nonstabilizing peptides). The 251 octapeptides with H-2K^b binding affinity are highlighted on the trained SOM presented with Figure 2(a). It is noteworthy that 241 of the 251 stabilizing peptides form a “stabilizing cluster” on the map (neurons 9–11/0, 9–10/1, 9–10/2), which indicates that these peptides are more similar to each other than to the randomly generated octapeptides. The highest density

of stabilizing peptides is located in neuron (9/1) which contains 180 sequences. The outlier neuron (1/15) contains octapeptide 9 (FHHAHRTV), a stabilizing octapeptide with a canonical residue in only one anchor position.

The distribution of stabilizing octapeptides fulfilling the canonical motif in all three anchor positions (80 sequences) is presented in Figure 2(b). All 80 octapeptides are located in a “stabilizing cluster”. Of the 152 sequences complying in only two anchor positions with the canonical motif 145 are located in this “stabilizing cluster” (Figure 2(c)). The remaining seven of the 152 sequences, which are not located in the “stabilizing cluster”, are located in neurons

framing the “stabilizing cluster”. The canonical motif is thus overrepresented in the “stabilizing cluster”. Although the known active octapeptide sequences constitute an island on the SOM implying similar physicochemical properties, our experimental results suggest that the canonical motif represents only a, albeit maybe dominant, fraction of the MHC I stabilizing sequences (Table 1, Seq. 10–21).

The SOM presented in Figures 3(a)–3(c) presents the distribution of sequences containing only stabilizing (green), only nonstabilizing (red), or containing both stabilizing and nonstabilizing octapeptides (blue). The locations of all 251 stabilizing octapeptides are shown; Figure 3(a) additionally includes category (i) nonstabilizing octapeptides, Figure 3(b) category (ii), and Figure 3(c) category (iii) peptides. The majority (54%) of the nonstabilizing octapeptides of category (i) (24 sequences) is located in neurons surrounding the “stabilizing cluster”, implying similarity in terms of the peptide representation by physicochemical properties (Figure 3(a)). Notably, the three neurons (9/1), (10/2), and (11/0) also contain nonstabilizing peptides (blue-colored neurons). Four motif-conform nonstabilizing octapeptides (WRYNYDPL, FRYEYRSL, HRYVYRNI, YRYKYDRL) are located in neuron (9/1) which contains the highest number of stabilizing octapeptides (180 sequences). The remaining (46%) nonstabilizing octapeptides of category (i) populate the lower left quadrant of the SOM. As illustrated in Figure 3(b), this area becomes more densely occupied when the 93 octapeptides of category (ii) are included: 75% of these peptides are located in this area of the SOM. Only two sequences of category (ii) can be found in the “stabilizing cluster”: FRYVWRTL and TTEWYTKI (neurons (9/0) and (9/1), Figure 3(b)). Apparently, category (iii) octapeptides are scattered in sequence space (Figure 3(c)). Only one nonstabilizing octapeptide of this category can be found in the “stabilizing cluster” (neuron (10/0), Figure 3(c)).

In summary, we have identified nine stabilizing octapeptides, two of which conform to the canonical motif in all three anchor positions, six fulfill two anchor requirements, and one sequence complies with the canonical motif in just one position. The majority of the designed and tested octapeptides (171 sequences) had no MHC I stabilizing effect. Twelve of the nonstabilizing octapeptides completely conform to the canonical motif, and sixteen fulfill the motif at two anchor positions. 23 octapeptides comply with the canonical motif at one residue position and exhibit no stabilizing effect. The remaining 120 octapeptides share no residue position with this motif. Since the experimental results reported here were not included in the SOM training, the resulting map provides a physicochemically defined distribution of stabilizing and nonstabilizing octapeptides in sequence space. Apparently, the stabilizing octapeptides constitute an island in octapeptide sequences space. Still, nonstabilizing octapeptides are collocated in this area. These nonstabilizing samples fulfill three or two residue positions of the canonical motif. The SOM clusters stabilizing peptides in a section of sequence space with similar physicochemical properties. A hint towards additional stabilizing clusters could be sequence 9 which is not clustered together with the other stabilizing peptides. Furthermore, neurons adjacent

to the “stabilizing cluster” also contain MHC I stabilizing sequences, which indicates that the epitope motif concept need to be extended in order to cover and predict alternative stabilizing peptides.

4. Conclusions

Our study confirms and extends the epitope motif concept for MHC-binding peptides proposed by Rammensee and coworkers [6]. We found octapeptides that lack key anchor residues but still exhibit a pronounced MHC I stabilization ability well comparable to peptides that fully conform to the canonical sequence motif. We also present a number of motif-conform but nonstabilizing peptides. This two findings clearly demonstrate that the canonical sequence motif alone is no sufficient criterion for the MHC I stabilizing peptides.

Acknowledgments

The authors are grateful to Norbert Dichter for technical assistance. This study was supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften and the Hermann-Willkomm-Stiftung, Frankfurt, Germany. J. M. Wisniewska and N. Jäger contributed equally to this study. G. Schneider and J. A. Hiss share senior authorship.

References

- [1] A. R. M. Townsend, J. Rothbard, and F. M. Gotch, “The epitopes of influenza nucleoprotein recognized by cytotoxic T lymphocytes can be defined with short synthetic peptides,” *Cell*, vol. 44, no. 6, pp. 959–968, 1986.
- [2] P. J. Bjorkman, M. A. Saper, and B. Samroui, “The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens,” *Nature*, vol. 329, no. 6139, pp. 512–518, 1987.
- [3] H.-G. Rammensee, K. Falk, and O. Rötzschke, “Peptides naturally presented by MHC class I molecules,” *Annual Review of Immunology*, vol. 11, pp. 213–244, 1993.
- [4] A. Sette, A. Vitiello, B. Rehman, et al., “The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes,” *The Journal of Immunology*, vol. 153, pp. 5586–5592, 1994.
- [5] R. C. Su and R. G. Miller, “Stability of surface H-2K^b, H-2D^b, and peptide-receptive H-2K^b on splenocytes,” *Journal of Immunology*, vol. 167, pp. 4869–4877, 2001.
- [6] H.-G. Rammensee, J. Bachmann, N. P. N. Emmerich, O. A. Bachor, and S. Stevanovic, “SYFPEITHI: database for MHC ligands and peptide motifs,” *Immunogenetics*, vol. 50, no. 3–4, pp. 213–219, 1999.
- [7] M. Dorigo and G. D. Caro, “Ant colony optimization: a new meta-heuristic,” in *Proceedings of the Congress on Evolutionary Computation*, pp. 1470–1477, 1999.
- [8] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, MIT Press, Cambridge, Mass, USA, 1986.
- [9] R. Brock, K.-H. Wiesmüller, G. Jung, and P. Walden, “Molecular basis for the recognition of two structurally different major histocompatibility complex/peptide complexes by a single T-cell receptor,” *Proceedings of the National Academy of Sciences of*

- the United States of America*, vol. 93, no. 23, pp. 13108–13113, 1996.
- [10] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of the IEEE International Conference on Neuronal Networks*, vol. 4, pp. 1942–1948, 1995.
- [11] M. Dorigo and T. Stützle, *Ant Colony Optimization*, MIT Press, Cambridge, Mass, USA, 2004.
- [12] M. J. Blythe, I. A. Doytchinova, and D. R. Flower, “JenPep: a database of quantitative functional peptide data for immunology,” *Bioinformatics*, vol. 18, no. 3, pp. 434–439, 2002.
- [13] P. A. Reche, H. Zhang, J.-P. Glutting, and E. L. Reinherz, “EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology,” *Bioinformatics*, vol. 21, no. 9, pp. 2140–2141, 2005.
- [14] B. Peters, J. Sidney, P. Bourne, et al., “The immune epitope database and analysis resource: from vision to blueprint,” *PLoS Biology*, vol. 3, article no. 99, 2005.
- [15] M. Bhasin, H. Singh, and G. P. S. Raghava, “MHCBN: a comprehensive database of MHC binding and non-binding peptides,” *Bioinformatics*, vol. 19, no. 5, pp. 665–666, 2003.
- [16] W. Zhong, P. A. Reche, C.-C. Lai, B. Reinhold, and E. L. Reinherz, “Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire,” *The Journal of Biological Chemistry*, vol. 278, no. 46, pp. 45135–45144, 2003.
- [17] J. A. Hiss, A. Bredenbeck, F. O. Losch, P. Wrede, P. Walden, and G. Schneider, “Design of MHC I stabilizing peptides by agent-based exploration of sequence space,” *Protein Engineering, Design and Selection*, vol. 20, no. 3, pp. 99–108, 2007.
- [18] N. Jäger, J. M. Wisniewska, J. A. Hiss, et al., “Attractors in sequence space: agent-based exploration of MHC I binding peptides,” *Molecular Informatics*, vol. 29, no. 1-2, pp. 65–74, 2010.
- [19] P. P. Grassé, “The automatic regulations of collective behavior of social insect and “stigmergy,”” *Journal de Psychologie Normale et Pathologique*, vol. 57, pp. 1–10, 1960.
- [20] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: from Natural to Artificial Systems*, Oxford University Press, New York, NY, USA, 1999.
- [21] P. Baldi and S. Brunak, *Bioinformatics: The Machine Learning Approach*, MIT Press, Cambridge, Mass, USA, 2nd edition, 2001.
- [22] H.-G. Ljunggren and K. Kärre, “Host resistance directed selectively against H-2 deficient lymphoma variants. Analysis of the mechanism,” *Journal of Experimental Medicine*, vol. 162, no. 6, pp. 1745–1759, 1985.
- [23] G. Köhler, K. Fischer-Lindahl, and C. Heusser, “Characterization of a monoclonal anti-H-2K^b antibody,” *The Immune System*, vol. 2, pp. 202–208, 1981.
- [24] R. Vita, L. Zarebski, J. A. Greenbaum, et al., “The immune epitope database 2.0,” *Nucleic Acids Research*, vol. 38, pp. D854–D862, 2010.
- [25] M. Nielsen, C. Lundegaard, P. Worning, et al., “Reliable prediction of T-cell epitopes using neural networks with novel sequence representations,” *Protein Science*, vol. 12, no. 5, pp. 1007–1017, 2003.
- [26] B. Peters, H.-H. Bui, S. Frankild, et al., “A community resource benchmarking predictions of peptide binding to MHC-I molecules,” *PLoS Computational Biology*, vol. 2, no. 6, pp. 574–584, 2006.
- [27] T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.
- [28] G. Schneider and P. Wrede, “Artificial neural networks for computer-based molecular design,” *Progress in Biophysics and Molecular Biology*, vol. 70, no. 3, pp. 175–222, 1998.
- [29] G. Schneider, M. Hartenfeller, M. Reutlinger, Y. Tanrikulu, E. Proschak, and P. Schneider, “Voyages to the (un)known: adaptive design of bioactive compounds,” *Trends in Biotechnology*, vol. 27, no. 1, pp. 18–26, 2009.
- [30] P. Schneider, Y. Tanrikulu, and G. Schneider, “Self-organizing maps in drug discovery: compound library design, Scaffold-Hopping, repurposing,” *Current Medicinal Chemistry*, vol. 16, no. 3, pp. 258–266, 2009.
- [31] K. Udaka, T. J. Tsomides, and H. N. Eisen, “A naturally occurring peptide recognized by alloreactive CD8⁺ cytotoxic T lymphocytes in association with a class I MHC protein,” *Cell*, vol. 69, no. 6, pp. 989–998, 1992.
- [32] K. Pearson, “Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia,” *Philosophical Transactions of the Royal Society of London*, pp. 253–318, 1896.
- [33] N. Auphan-Anezin, C. Mazza, A. Guimezanes, et al., “Distinct orientation of the alloreactive monoclonal CD8 T cell activation program by three different peptide/MHC complexes,” *European Journal of Immunology*, vol. 36, no. 7, pp. 1856–1866, 2006.
- [34] D. R. Madden, “The three-dimensional structure of peptide-MHC complexes,” *Annual Review of Immunology*, vol. 13, pp. 587–622, 1995.